DOCUMENT RESUME

ED 390 252 FL 023 178

AUTHOR Shiozawa, Tadashi; Simmons, Thomas
TITLE Do Improved Examination Scores Indicate Acquired Skills? A Study in Teaching the Use of Articles in Journal Writing through a Continuous Testing Regimen.
PUB DATE 95
NOTE 19p.
PUB TYPE Reports - Research/Technical (143) -- Journal Articles (080)
JOURNAL CIT Journal of International Studies; n14 p133-149 1995

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Determiners (Languages); Educational Strategies; *English (Second Language); Foreign Countries; Grammatical Acceptability; *Journal Writing; Language Research; Language Usage; Scores; Secondary Education; Second Language Instruction; *Testing Programs; *Test Interpretation; *Test Validity
IDENTIFIERS Japan

ABSTRACT

A study undertaken in Japan investigated the degree to which students' use of English as a Second Language in journal writing outside of class corresponded to their performance on language tests. Specifically, the correct use of articles in the position and function of "a, an, and the" was examined. Subjects were students in a secondary school specializing in language training. The experimental group, 132 students in 6 second-semester classes, was instructed in the use of articles; the control group of 222 students in 6 first-semester classes, was not. Student journal extracts were randomly sampled to measure use of articles, and results were compared with quiz scores. It was found that neither group showed any significant improvement in use of articles in journal writing; however, the scores on quizzes in the experimental group consistently increased as they were given instruction on article use. Results suggest that while the students were gaining knowledge from instruction, they were not applying it in practice. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

To: Center for Applied Linguistics
1118 22nd Street NW
Washington, DC 20037-0037

ED 390 252

Do Improved Examination Scores Indicate Acquired Skills? – A Study in Teaching the Use of Articles in Journal Writing Through a Continuous Testing Regimen –

Tadashi SHIOZAWA
Thomas SIMMONS

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Thomas Lee Simmons
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

中部大学
国際関係学部
紀要No.1抜刷
1995・3
Journal of International Studies
Chubu University, Kasugai, JAPAN

FL023178

BEST COPY AVAILABLE

Do Improved Examination Scores Indicate Acquired Skills? – A Study in Teaching the Use of Articles in Journal Writing Through a Continuous Testing Regimen –

Tadashi SHIOZAWA*
Thomas SIMMONS**

I. Introduction

For the last 6 years the authors have been investigating mastery learning as a goal in our teaching to establish some gains that the students can take with them when they leave schools (Simmons, 1992, Shiozawa 1992). We have observed that discrete variables are comparatively easier to teach in the test-retest approach (Dempster, 1992, pp. 213-217) because they are retained better by the students. Additionally, discrete variables are easier to manage because they can be dealt with in short, focused cloze tests or multiple choice tests that can be discussed quickly when the students have finished. However, while we have been able to verify the efficacy of the test-retest approach in improving the students' scores and its usefulness in managing the programme's heavy schedule and large class loads, the question remains as to whether the students are effectively incorporating the tested objectives into their writing assignments. An initial study of the student's journals written during the previous year was conducted after they had been tested and retested for the use of deictic variables in writing but had not been given any specific directions to incorporate what they were learning into their journals which were being written at the same time. The preliminary survey indicated that they were incorporating the objectives that were being taught, but that they still had some ways to go before they had mastered the use of the objectives. It was the purpose of this study to establish whether or not the students can be given an advantage in learning and employing the objectives by examining the possible influence that using the continuous testing has on their use of the tested variables in their writing. In other words, we set out to learn whether or not the students incorporate the discrete variables taught through a continuous testing approach into their writing assignments.

* 外国語教室 (Department of Foreign Languages)
** 日本大学 (Japan University)

Ⅱ. Institutional parameters and the rationale for this study

One of the difficulties faced in conducting this research was that the syllabus assigned to the teachers at the school wherein this study took place[1] included discrete objectives that were part of the mandatory learning objectives. Administrative requirements aside, the proper use of these discrete objectives is also considered important in oral and written communication since they are noticeable when used incorrectly and tend to be annoying and frequently the cause of ambiguity. But they are tedious to study at length and the students are often distracted if the objectives are given more than 10 to 15 minutes during class; an attitude that is a further detriment to the classroom environment. Reducing the time spent on such objectives could possibly decrease the students' lack of enthusiasm for learning them and arguably boost teacher morale.

Additionally, in the course of teaching large classes it becomes necessary to deal efficiently and quickly with an overwhelming number of classroom management variables. Thus, it is expedient to deal with discrete objectives quickly and consistently so that the students will have an opportunity to master them and the teacher will be able to handle them in the inadequate time allowed for an accumulating number of objectives through the course of the year. So, it is important to find straightforward ways of dealing with discrete objectives in a well delineated format that can be taught with a minimum of preparation using a teachers' resource bank, evaluated quickly and produce consistent results. Such an arrangement will significantly reduce the classroom teachers' workload, thereby releasing teachers for other more complex or esoteric aspects of teaching writing.

Ⅲ. Hypothesis

Using repeated tests at frequent intervals to promote learning (Dempster, 1992, pp. 213-217), we have been able to see improved exam scores and smaller standard deviations throughout the entire student population as the higher level students reach mastery quickly and the lower level students catch up. However, we were unable to see how the students used their skills outside of class and thus could not determine if they were incorporating the lessons from class into their use outside of class – the real goal of teaching communication in English. The only way the students were consistently using the English skills they learnt in class (with the help of the repeated tests and other means) had been the journals which were written at all hours during and in between the

school sessions. The journals, then, provided the only way to measure their actual use of the language as they employed it outside of the classroom environment and was probably the closest thing possible to their unmonitored and unsupervised use of the language. Therefore, it is reasonable to hypothesize that if the students are using their skills learned in class outside of class, an analysis of their journals may show a significant increase in the students' correct use of the articles. If not, the analysis of their journals should show that the average scores of the definite and indefinite articles correctly used would not increase significantly whereas the scores of the short quizzes on articles given to the subjects inside class should increase as the term goes by.

We can also predict that those students who are already at a fairly high level of competence in the use of articles would probably not show significant gains in their use of the articles, but neither should they regress, the students who are seriously deficient in their use of the articles and similar determiners should show a steady improvement in the exams, and may show improvement in their use of the determiners, and the survey should show that the standard deviations for the lower groups compared to the upper groups and the intra-group scores of the lower level students evinced a smaller and smaller range as the total population reached parity in their use of the articles.

Ⅳ. Methods

A. Research Design

Finding an efficient use of the teachers' time in the course of a school year is a priority in large class situations. This is even more pronounced in writing classes that must employ a heavy production load that is to be monitored and evaluated by the teacher. There exists the technique of continuous testing that has shown significant gains in students' test scores, but writing teachers (as do any other professional educators) must also concern themselves with developing the students' ability to communicate in the written medium of expression. In a study of 132 students in a post-secondary vocational school in Tokyo Japan, a continuous testing regimen employing short quizzes and a short discussion of the results was employed exclusively to instruct the students in the use of definite and indefinite articles and the students' actual writing assignments were then surveyed to ascertain their proficiency in the use of articles and related determiners.

We chose the discrete objectives of definite and indefinite articles and related determiners as the independent variables (Ary, Jacobs, Razavieh, 1990, p. 31) in the test

retest intervention because they are particularly important for EFL in Japan: They are more easily defined than other objectives; they are arguably an important function of precise communication in English; they have no corollary in the Japanese language; and Japanese tend to neglect their use in speaking and writing.

The number of times the students used the articles and other corresponding determiners correctly and incorrectly in their writing assignments were tabulated. The students' scores were then recorded and statistically analyzed. The criteria of evaluation are as follows.

Table 1

Table of all specific determiners (Halliday and Hasan, 1976: 70)

	Demonstrative	Possessive
Referential	Selective *this, that* *these, those*	Speech roles *my, your, our*
		Other *his, her, their* *its*
	Non-selective *the*	*one's*
Interrogative	*which*	*whose*

The students' randomly selected journal writing samples were evaluated with regard to determiners that correspond to the position and function of the articles 'a', 'an' and 'the'. Those that co-occur with nouns in the article position and do not act in the position of the head were considered in the students' usage with the following provisions (see table 1):

1. Adverbial (circumstantial) demonstratives, (here, there, now, then , etc., as defined by Halliday & Hasan, 1976, pp. 57-58), were not included:

a) The textbook does not address the location of a process in space or time except to indicate if a noun is definite (particular) or indefinite (general) or has been mentioned before in the text (new or old information);

b) They are typically used as adjuncts in a clause and not as elements in the nominal group that precede the noun, they would not demonstrably provide the students a direct means of associating their similarities;

2. The demonstrative references were counted when used properly because they are easily used in place of the articles and also carry the meaning of plural and singular

nouns (this, that, these, those).

3. Possessive determiners that act as modifiers, my, your, our, his, her, their, its, one's (delineated by Halliday & Hasan, 1976, pp. 38, 45, 54) were initially counted because they occupy the article position (my part time job → the part time job), but they were excluded from the study because the students showed a clear mastery of their use.

4. Non-specific deictics (each, every, any, either, neither, no, some, all, and both) with the exception of the indefinite article 'a' were all initially counted except when they functioned as the head of an elliptical nominal (Halliday & Hasan, 1976, pp. 100-101, 157-59), but they too were excluded when it became clear that both student groups evinced no significant difficulty in their use.

5. The definite article may operate exophorically when there are examples of common use (I went to the hospital) and such use was counted as correct, otherwise the students were expected to establish new information with 'a' and then proceed to use the definite article (Halliday & Hasan, 1976, p. 70).

B. Procedure

To check the reliability of our scoring procedure, a random sample of ten journal extracts from each group were photocopied and then marked. Three weeks later the same journal extracts were again photocopied and marked and then the second marking was compared to the first marking. These two tests were averaged together. The agreement between the two markings indicated an acceptable[2] degree of reliability in the evaluation of the students' journals.

Table 2

Score Reliability

			1st	2nd
mean	98%	agreement	73.4	72.15
sdev	94%	agreement	13.84	13.07

The subjects in the experimental group were in six different classes for a total of two-hundred and forty. These students were in their second semester of the 93-94 year class, taking an introductory level composition class at a post-secondary vocational school that specialises in teaching foreign languages for entry into the translation, interpretation, tour guide, and travel agency businesses. Only one-hundred and thirty-two students of the total group were measured because the other one hundred and eight did

not meet basic compliance guidelines.

The subjects in the initial study group were in six classes of two hundred and forty-eight students in their first semester in the 93-94 class of the same institution with the same curriculum. Only two hundred and twenty two students of the total group were measured because the other twenty-six did not attend enough classes or did not write journals that would have given them the opportunity to use what they learned in class.

The experimental subjects took a weekly series of short quizzes specifically constructed to incorporate textbook objectives in the use of determiners (specifically articles), whereas the students in the control group did not. The quizzes were fill in the blank that focused entirely on the determiners, given at regular intervals in each class and the results were discussed with them for a total of ten to fifteen minutes per class (see appendix). The quizzes were tabulated to ascertain whether the students were showing any improvement in scores. The one-hundred and eight subjects not included in the study were excluded because they did not take at least 50% of the tests and/or show evidence of having done any writing which would have presented the opportunity to use what they had learned.

In addition the subjects also had a series of letter writing assignments and maintained journals throughout the semester. Since the groups in the study were assigned to their classes on the basis of their results on The Oxford Placement Exam, they have, in effect, been stratified. For this reason, we chose to sample a proportional number of students from each level to reflect the proportion in their classes (Ary, Jacobs, Razavieh, 1990: 173-75). The student's journals in both groups were randomly sampled to measure their use of the articles and other determiners. Twenty four students four from each class, were chosen to be examined in both the experimental and control groups. The use of the determiners was classfied as:

Used correctly	Used incorrectly	Incorrectly omitted
yes	*no*	*O*

Data was evaluated for the correct use (y), incorrect use (n) or incorrect omission (o) of articles and related determiners[3].

It should be noted here, however, that this experiment could only qualify as a pre-experiment (Nunan, 1992, pp. 40-41; Larsen-Freeman, Long, 1991, pp. 19-21) for the following reasons.

a) There were pre- and post-tests' in the form of the journal surveys taken at the beginning and the middle of the semester;

b) There was no control group during the observation and the group surveyed initially did not qualify as a control group. Although it was composed of 1st semester students in the same class, 93-94, they were neither chosen randomly nor were they at a similar maturation level in their language studies. Additionally, the students' journal compositions were surveyed at different times in the course of the semester. Both the initial and experimental groups' writing samples were taken at the beginning of the semester (prior to the testing regimen for the experimental group) and were analyzed for a baseline study. However the initial groups' writing was sampled again at the end of the 1st semester (after nearly 5 months in class) and the experimental groups writing (surveyed after the testing regimen) was sampled near the middle of the second semester — after nearly 8 months of instruction;

c) There was no random assignment of class placement as such. The students were placed in their respective classes on the basis of a perceived difference denoted by the 'Oxford Placement Test' (OPT) and the students themselves had selected their study major which also determined the class in which they were placed. The students' course of study did not differ from the students in other courses with regard to their foreign language studies, merely with regard to their vocational studies. However, with the exception of the possible influence of the factors mentioned, all of the students in the measured selection were assigned to classes in this study in such a way that all the levels of competence according to the OPT were included in the initial and experimental groups; there was a reasonable cross section of the populations' level of ability.

V. Results

Table 3 and 4 summarizes the results of the experiments. Table 3 shows the results of the scores tabulated from the journals twenty-four students wrote at the beginning and the end of the semester. Table 3 shows the mean scores of the journal survey and the percentage of the articles used correctly. A T-test was conducted to examine whether their increases are statistically significant. Table 4 gives the same information for the controlled group.

There was a 7.6% increase in the controlled study groups' sample mean (65 → 72.6%) and a 8.0% increase in the standard deviation. However, neither increase was statistically significant (t = 1.04 n. s.). The range of the population mean was calculated

Table 3

Journal survey results for the experimental group

	Beginning of Semester					Mid semester				
criteria	y	n	O	total	%	y	n	O	total	%
mean	26.3	1.63	6	33.4	78	27.2	1.087	5.38	33.7	78.7
sdev	11.5	2.02	4.02	13.3	10.2	12.9	1.21	4.23	13.6	14.8
t test: mean 1&2	0.67	0.16	0.53	0.95	0.81					
	n. s.	n. s.	n. s.	n. s.	n. s.					
Range: pop. mean	73.7 to 82.3					72.5 to 85				
df 23										
N = 132										
Range of scores	57% to 98%					32% to 100%				

* = $p < .05$, n. s. = non significant

Table 4

Journal survey resuts for the controlled group

	Beginning of Semester					Mid semester				
criteria	y	n	O	total	%	y	n	O	total	%
mean	12.0	1.3	3.7	17.0	65.0	10.0	1.2	3.5	14.0	72.6
sdev	7.9	1.8	3.8	10.0	17.6	7.4	3.1	3.5	10.0	25.6
t test: mean 1&2					1.04					
					N. S.					
Range: pop. mean	53.0 to 77.0					61.0 to 84.0				
df 23										
N = 222										
Range of scores	0% to 100%					0% to 100%				

* = $p < .05$, n. s. = non significant

(after Wonnacott & Wonnacott, 1985, pp. 228-229) and there was an increase in the range (53 → 77% to 61 → 84%). By the same token, the experimental group evinced no significant increase in their sample mean (78 → 78.7%) although there was an increase by 4.6 % in the standard deviation (10.2 → 14.8%). There was also a slight change in the population range (73.7 → 82.3% to 72.5 → 85%). However, the result of the T-test shows that the increase in the overall mean scores is statistically non-significant for this group as well ($p > .05$).

The lower ability groups' mean[4] ($n = 9$) in the journal survey declined somewhat (74.8 → 71%) and their standard deviation increased (9.7 → 17.45%). On the other hand, the higher groups' mean ($n = 8$) in the journal survey slightly increased (79.3 → 81.7), but not statistically significant enough ($p = .59$, n.s.).

The average scores of the weekly short quizzes given to the students in the experimental group increased by 10.1 % (67.6 → 77.6) . The mean scores of each short quiz were 67.6, 68.8, 74.6, 69.6, 76.0, 77.7, which showed a steady increase as the semester went by. The drop between the third and fourth mean scores corresponds to the student's vacation where in the score after vacation was lower.

To sum up the results, neither experimental nor controlled group showed any significant increase in the scores of the articles used correctly in journal writing. However, the scores of the short quizzes given to the experimental group have consistently increased as they were given instructions and reviews on articles in class.

VI. Discussion

Possessive determiners (my, our, his, etc.) and selective demonstratives (this, that, etc.) have complements in Japanese and apparently are not particularly difficult for the students to incorporate in their writing in English. During the course of tabulating the results, the possessive determiners were eliminated from the analysis. The possessive determiners were used frequently but the total population of both the initial study group and the experimental group made a sum total of 13 mistakes in their use of the possessives and 5 of those mistakes were made by one student. The selective determiners, on the other hand, were used infrequently and there were no mistakes for the entire population. They were nevertheless counted in the tabulation since they could be used interchangeably by the students. The non-specific deictics (each, every, any, etc.) were also discounted since they were used incorrectly only eight times in the entire survey.

The vast majority of mistakes were in the omission of the students' use of the

articles rather than their incorrect use. This would seem to support the view that they do not use them since they have no correlating words in Japanese. This failure to use the articles at all rather than using them incorrectly was a revelation. It indicates that there might be the need to approach the teaching of the articles through the use of the students' texts rather than those prescribed in the textbook because the situations wherein they use the articles are not sufficiently covered by the textbook.

There were no significant changes in any of the measured parameters for either group. There was a small change in the number of misused articles but the incorrect omissions were unchanged. Possible explanations for this lack of change, were not addressed by the study.

The students in the initial survey showed marginal changes – more so than the experimental group. These changes initially led us to believe that their progress could be accelerated. However if the differences between the beginning point of the initial survey are compared with the end point of the experimental survey, the change is still insignificant.

The experimental group showed very little increase in the journal survey results (78% → 78.7) despite the fact they were consistently given the reinforcement by the weekly short quizzes. The control group showed some improvement (65.1 → 72.6%), but this improvement is not a significant one. (p=1.04) The results of the quizzes they took , however, showed a steady increase (67.6 →77.6%). These two facts seem to reveal an extremely interesting finding, i.e. they were not using what they learned in class in the journal writing, a real out-of-class communication medium. They were learning determiners in class for the in-class use, mostly to improve the scores of the next quiz, not for the purpose of improving their communication skills in English. This suggests that what language learners can easily do in class may not be as easy as when they face a real communication situation outside the classroom.

One incident illustrates the above point well. One of the authors took Japanese college students to a US college for a short-term stay. He noted that most of the students, whose English was at the intermediate level, had trouble in a communication task as easy as introducing themselves when they actually had to talk to their American roommate on the first day of their stay. They completely forgot to say, " Hello." My name is so and so. Nice to meet you." When they were greeted with "How are you doing?" from their roommate for the first time, they couldn't possibly imagine the expression was a variation of "How are you?", one of the easiest and first expressions they learned when they started to study English at 12 years of age. Some of the participants on this prog-

ram expressed the view that they had never thought their English was as bad as it was. Thus, applying what they learned in a 'safe' classroom environment to an actual communication is not as easy a task as language teachers may imagine.

Pedagogically, the above fact also suggests that the language teachers may need to employ the communicative approach in class in a more focused manner. Giving the learners as many chances to use the language in a meaningful way as possible seems to be a key element in light of our findings. The teachers may also need to give the learners the chances to experience a variety of "Englishes" and a variety of situations. For example, they may need to have their students listen to English with some accent, with some noise, under some mental pressure, or beyond their students' comprehension level. Quizzes or regular mid-term or final exams may also need to be more proficiency oriented rather than achievement oriented ones if they believe the real purpose of teaching English is for the learners to use the language when they need to in the outside world.

The most important finding from the above mentioned facts is probably about using test scores for measuring students' improvements of their language competence. Our findings suggest that we may not be able to, or sometimes should not, use paper-and-a-pencil tests for measuring affects of any language teaching methods on the competence of the students. This is because they may do well in the tests trying to measure their improvement of the target language items, but they may not be able to apply them to outside class language use, which researchers are supposedly interested in measuring. The language teachers and researchers may need to keep it in mind that what their students show in the test performance may be completely different from what we may call "communicative competence". In other words, scores of the tests given in language classroom may mean something only inside the classroom, not outside.

The initial conjecture that the lower level students would show the most change can not be supported in light of these results but the overall level of writing may have changed. In neither case was the level of language investigated – the students may have actually made significant improvement in the language skills to a level where the use of determiners and articles are more complex and therefore more difficult. But the study would not have shown this or taken it into account.

The lower ability students' group mean in the journal survey declined (74.8 → 71.0%) and the standard deviation for the lower level students did not get smaller, which is opposite from what we had predicted. It in fact increased from 9.7 to 17.5. This may mean that the lower level students had more difficulty utilizing what they had learned in class in actual journal writing than upper level students. Some low level stu-

dents may have got more confused with the use of articles as the instructions on articles became more complex towards the middle of the semester and could not use what they have learned in the actual journal writing, whereas other low level students started to apply the information they have acquired in class to journal writing. To sum up, it seems the lower level students tend to fail to use the information learned in class in actual communication more than upper level students do.

The high ability students' mean score in journal writing did not increase either, which supports our prediction. This suggests the difficulty in acquiring accurate knowledge in article use even for high ability students. This is most probably due to the nature of the Japanese language, which lacks the concept of articles, but this cannot be proved by this study. The fact that even high level students had difficulty in transfering language information to language skills means that it was not because of the level of proficiency of the language that the subjects failed outside class to use what they learned in class. Therefore, it can be concluded that language learners in general would have difficulty in transfering in class this specific language information to outside class language use.

The last thing we wanted to test was whether the in class continuous testing regimen on articles improved the accuracy and the fluency of the articles used in the journal writing. Judging from table 3 and 4, something we may infer is that the short quizzes did not have significant effect on the scores on the journal writing in the experimental group (78.0 → 78.7) (t = .67, n.s). On the other hand, surprisingly enough, the average scores of correctly used articles for the journal writing in controlled group improved more drastically than that in experimental group (65.0 → 72.6). However, the increase is not yet statistically significant (t = 1.04, n.s). We do not know exactly why the control group without short weekly quizzes have shown more improvement. One possible explanation is that those who were in the controlled group could study articles in a more relaxed manner because they did not have to take quizzes so often. They could have practiced articles in class with a long term objective in mind, which is to use them in the journal writing, whereas those in the experimental group worried too much to improve the quiz scores; thus, failed to use them correctly when they had to use them for communication. When the quizzes were over, they might have forgotten all about what they have learned in class. At any rate, it seems that the numerous weekly short quizzes served only to make the students study for the quizzes themselves, not to improve communications skills using articles.

Ⅶ. Shortcomings and suggestions

Further study into these areas are certainly indicated. What is needed is a valid method of evaluating the students' overall writing skills and a thorough analysis of the precise way in which they use the articles and related determiners.

Another difficulty was that during the time of the study, the teachers themselves carried student loads that usually exceeded 10 classes a week. Because the large class loads (40+ students) and inadequate time (80 minutes once a week) in class prevented the teachers from consistently dealing with the problems and deficiencies they saw first hand and made the goal of mastery learning problematic. For this reason, this study was undertaken in the hope that it would facilitate the development of a workable approach that would be manageable for the teachers within the restrictive parameters so common in foreign language education in post-secondary vocational colleges in Japan (Shiozawa, T., Simmons, T., & Noda, K., 1993; Shiozawa, T. & Simmons, T., 1994).

The quasi-experimental design (Seliger and Shohamy, 1989, pp. 148-149) offers a natural extension of this study because there were control groups available at the institution. All students have had a standardized pre-test that would have allowed comparisons with other students in other classes that were not subjected to the independent variable. When the teaching schedule ended at the end of the semester, those groups could have been delineated and studied and the survey of the students in the experimental group could also have been conducted again at the end of the semester. The intermediate evaluation that this study represents could enable the investigator to use more accurate data in developing the tests themselves. For these reasons the study design can be used to accommodate a second evaluation.

Ⅷ. Conclusion

The students' test scores improved almost as expected. The parameters outlined by Dempster were easily duplicated with objective criteria in this survey and the students behaved almost as predicted. However, even though the students at the bottom end of the class overcame the 75%+ score range on their tests they did not show any improvement in their use of the articles in journal writing – they were apparently not transferring their knowledge to their skills. The students who scored well on the tests did not show much improvement but then it is difficult to show much relative improvement above 75%. So the test scores for the higher students could not really show signifi-

cant improvement.

The hypothesis that the lower level students (below the 75% range on the tests) would show improvement in skill can not be born out because the lack of significant gains as indicted by the journal survey. The assumption that test scores indicate the improvement of their skill in using 'the' articles was unfounded. This indicates the need to be wary of any similar assumptions in the classroom. This also points to the teachers' limitations. In so far as educators are supported in their attempts develop the opportunities to directly relate the educational objectives is the extent to which they can be actively involved in teaching the very skills targeted. Using tests, at least in this area of knowledge, does not adequately evaluate the students' knowledge and skill.

As stated earlier, the use of continuous testing as a large class management technique is effective in that it allows the teacher to monitor and evaluate the class for administrative purposes. However, this approach, which has been successfully duplicated numerous times, merely points to the deficiencies in the educational environment faced by students and teachers. In other words, even though exam scores can be demonstrably improved, this does not mean the students have actually learned the targeted skill of using the information; they are only improving their exam scores.

It is obvious that it takes time for language learners to develop language information into language skills, the competence they can actually use in communication. For example, they may be able to pronounce a certain target pronunciation item one day, but that does not guarantee that they could do the same thing with the same ease the next day or that they could pronounce it correctly in an actual communication with native English speakers. Our research highlighted this simple fact. If it serves to remind language teachers to be a little patient in strenuous teaching conditions experienced every day and warn language researchers to be extra careful in interpreting test results given in class, this study, which took so much time and energy of our students and ourselves, was worth conducting .

Notes

[1] Japan College of Foreign Languages (Nihon Gaikokugo Senmon Gako) in Tokyo, Japan, a subsidiary of the Bunsai Gakuen Foundation in Osaka.

[2] The process of rating intra-marker reliability is similar to that used in statistically evaluating reliability in reading radiographs. Comments are made and one month later, without an reference to the previous marks and comments, the process is done again and then the two are compared.

[3] Not incorporating correct omissions into the measured parameters might have revealed some

interesting data but the logistical load of demonstrating those measured parameters and then actually measuring them is something that could arguably only be done with such a large sample through the aid of sophisticated software beyond the resources of this study. It would, for example, included the counting of all nouns and the correct omission of the articles.

[4] Students whose test grades were in the bottom 1/3 of the class.

[5] Given the genre which was focused on the students personal experiences in what were predominantly narratives, the frequent use of possessives is not surprising (K. Marriage, Aston University, personal communication, 3/94)

References

Ary, D., Jacobs, L. C., & Razavieh, A. (1990). *Introduction to Research in Education*. Chicago: Holt, Rinehart and Winston, Inc.

Dempster, F. N. (1992). Using tests to promote learning: A neglected classroom resource. *The Journal of Research and Development in Education, 25* (4).

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. New York: Longman.

Larsen-Freeman, D. & Long, M. H. (1991). *An Introduction to Second Language Acquisition Research*. Essex: Longman Group.

Nunan, D. (1992). *Research Methods in Language Learning*. Cambridge: Cambridge University Press.

Seliger, H. W. & Shohamy, E. (1989). *Second Language Research Methods*. Oxford: Oxford University Press.

Shiozawa, T. (1992). Why can't high school graduates Write? *Journal of the College of International Studies, 8* . Aichi-ken, Japan: Chubu University.

Shiozawa, T, Simmons, T. & Noda, K. (1994). Origins and responses to conflict in the teaching environment. *The Language Teacher*. Vol. 18, No.4.

Shiozawa, T, Simmons, T. & Noda, K. (1993). Working conditions and career parameters in the educational environment for foreign teachers of languages in Japan: Conflict and resolution, discrimination, and empowerment. *Journal of the College of International Studies, 10*. Aichi-ken, Japan: Chubu University.

Shiozawa, T. & Simmons, T. (1993). Social and administrative parameters in methodological innovation and implementation in post secondary language schools in Japan. *Journal of the College of International Studies, 11*. Aichi-ken, Japan: Chubu University.

Simmons,T. (1992). What entails user-friendly tests of English as a foreign language? Paper read at Hakuba Linguistics Circle.

Wonnacott, R.& Wonnacott, T. (1985). *Introductory Statistics* (4th ed.). New York: J. Wiley & Sons.

Woods, A., Fletcher, P. & Hughes, A. (1986). *Statistics in Language Studies*. Cambridge: Cambridge University Press.

Appendix

A. textbooks

The textbook used during the study incorporated a modest section on the use of determiners which has proven inadequate for the students' use in their own writing and in addressing the questions they ask about the use of the determiners.

A selection was made from a writing specimen of the selected students and the proceeding categories were ascertained according to the following guidelines (*Word for Word* ", p. 32)

1. Person, Country, City:
Most names of things and people do not take an article

2. Particular (definite), General (indefinite):
Particular things often will take definite articles

3. Countable, Uncountable:
Uncountable nouns when used in a general sense do not have an article

4. Singular, Plural:
Singular, countable nouns must have an article or a word such as ***my***, ***this***, and so on. Plural nouns when use in a general sense have no article.

5. New, Old:
New information often has an indefinite article but things that have been mentioned before often have a definite article

B. Test Samples Used in the Testing Regimen (Taken Verbatim from the Text Used in the Classroom)

1. Word for Word Unit 7
Writing test for subjects covered on page 32-34 #2.2A
Use ***a***, ***an***, ***the*** or 'X' if none is needed.

My family went to Arizona in 1961 and we visited Grand Canyon. I had never seen Canyon. It was very beautiful. We stayed at hotel while we visited Canyon. Early in morning, I would leave the hotel and I would go to bottom of Canyon on horseback. I always rode same horse. horse was very big and gentle. It is very pleasant memory for me.

2. Word for Word Unit 7
Writing test for subjects covered on page 32-34 #2.2B
Use ***a***, ***an***, ***the*** or 'X' if none is needed.

Teresa is friend of mine. She met star of "Dances with Wolves", Kevin Costner. She is student of Amerindian cultures and wanted to ask him about Lakota and Cheyenne people who were in movie. When she saw movie, she was surprised that it seemed so authentic. Usually movies about Amerindian peoples are very strange and untrue. She says that Costner enjoyed making movie and has since made many Amerindian friends. He even speaks little Mandan and Kiowa lan

guage.

The following test was taken directly from the text book. The answers to the test were given at the back of the book. The test itself was given at the end of the third month while the unit itself was taught at the beginning of the first month in the semester.

Unit 7 2.3 Work with a partner to decide which article is needed where you see ___ in the following text. You do not need an article in every blank.

The Lion

Male lions are larger than females and have thick manes on their necks and shoulders. ___ lions rest for up to twenty hours a day and live in ___ group called ___ pride. ___ lioness does most of ___ hunting and is able to catch large animals like antelopes and zebras. Lions can have ___ cub in any season. Between one and six young can be born at ___ time.

6. Word for Word Unit 7
Writing test for subjects covered on page 32-34 #2.2E
Use *a, an, the* or 'X' if none is needed.

___ students at ___ Kofu West High School have read ___ book that is about ___ animals that live in ___ North America. After the students read ___ book, they saw ___ movie about ___ herd of bison in ___ Colorado Mountains. ___ bison herd is safe from ___ hunters and there are more bison each year. ___ hunters used to hunt them without any control and ___ herds of ___ bison began to disappear. Now ___ hunters can not hunt unless they ask ___ permission and pay ___ money.
(correct answers (ea) 17)